Springboard--Data Science Career Program
Capstone Project #2 - Yelp Sentiment Analysis
Project Proposal
By Kevin Cole
June 2020

Problem: Positive yelp reviews are crucial for a business. Yelp not only drives traffic to business but easily helps users make decisions about which establishment to visit. These decisions are often made based on reviews.

A restaurant owner may want to tailor the customer experience to items which lead to high reviews and make business decisions to eliminate/reduce anything that can lead to negative reviews.

The goal will be to use data science methods that will help establish a connection between the rating levels and what might drive them.

Clients: The target clients would be restaurant/business owners who are looking to improve their ratings. They would be interested in business decisions that could directly lead to a better customer review.

Data: The data is provided by Yelp for as an open dataset (https://www.yelp.com/dataset). This subset of yelp data is a 10.5 GB download containing 5 different JSON encoded files. These files contain data on the business, the customer review, the user, the datetime of visit, and additional tips left by the user.

Approach: A supervised approach will be used, predicting which category (good/neutral/bad) each review. Classification algorithms will be explored to find what features drive each category.

In addition, an unsupervised approach will be explored with Latent Dirichlet Allocation to determine what topics are associated with good/bad reviews and to what extent those topics are correlated with their rating.

Deliverables: The project deliverables will include Jupyter notebooks containing methods used and code to support the analysis, a final report detailing methods, conclusions, and recommendations, as well as a presentation slide deck.